Assignment 5 – Object Detection (MNISTDD-RGB)

Overview

In this assignment, you will implement and analyze a **YOLO (You Only Look Once)** object detector on the **MNIST Double Digits RGB (MNISTDD-RGB)** dataset. The dataset consists of 64×64 RGB images containing two digits (0–9) placed randomly with varied backgrounds and noise. Your goal is to **detect both digits** and **localize them using bounding boxes**, without segmentation.

You will adapt a **single-stage YOLO detector** (e.g., YOLOv5, YOLOv8, or a lightweight variant) to perform object detection on this dataset and evaluate your trained model using **Intersection over Union (IoU)** and **mean Average Precision (mAP)** metrics.

Model Size Constraint: Your final model (weights and architecture) must not exceed **200 MB**. Submissions exceeding this limit or unable to run on Colab GPU will not receive credit.

Dataset & Environment

- **Dataset:** MNISTDD-RGB (64×64 RGB images, two digits per image).
- **Splits:** 55K train / 5K validation / 10K test (test data will remain hidden and will be released after the submission deadline for benchmarking and discussion).
- **Device:** GPU (Colab or local).
- **Libraries:** PyTorch + Ultralytics YOLOv5/YOLOv8 or equivalent implementation.
- **Output:** Bounding boxes and class labels (no segmentation masks).

Part A — In-Lab (2% of course grade)

Goal: Implement and train a compact YOLO model using a subset of the MNISTDD-RGB dataset and demonstrate correct detection of digits.

A.1 Tasks

1. **Dataset Preparation:** Load and visualize a small subset (∼10K images). Prepare labels and bounding boxes in YOLO format.
2. **Model Setup:** Implement or adapt a lightweight YOLO model (e.g., `yolov5n` or `yolov8n`). Use pretrained weights if available.
3. **Training:** Train for 10–20 epochs on the subset. Track loss, IoU, and sample predictions.
4. **Validation:** Demonstrate detections with bounding boxes on validation images.

A.2 Expected Performance (in-class)

Baseline target: average **IoU** $\geq$ **0.7** on the validation subset after training. Implement your own function to compute IoU.

Part B — Take-Home (3% of course grade)

Goal: Train a full YOLO detector on the entire MNISTDD-RGB dataset, evaluate quantitatively, and analyze results.

B.1 Tasks

1. **Full Training:** Train on all 55K training images for sufficient epochs. Tune learning rate, confidence, and NMS thresholds.
2. **Evaluation:** Compute average IoU, precision, recall, and mAP on the validation set throughout training. Present detection examples with bounding boxes and confidence scores.
3. **Report:**
   - Clear reasoning on speed vs accuracy
   - Short discussion on YOLO performance vs two-stage detectors
   - Motivate your choice for learning rate, confidence, and NMS thresholds

Deliverables Checklist

- Notebook(s): `A5.ipynb` (runs top-to-bottom on a clean runtime).
- Report (PDF, at most 2 pages): includes metrics (IoU, mAP), plots, visualizations, and a brief discussion.
- Trained model weights (`.pt` or `.pth`) not exceeding **200 MB**.

Marking Breakdown (5% total)

Part A (In-Lab — 2%)

#	**Requirement**	**Value (%)**	**Notes (Expectations for Excellent)**
1	Dataset loaded and visualized correctly	0.5	Sample visualization showing bounding boxes and classes
2	YOLO model training	0.5	Loss and other metrics are tracked (you may use implemented trackers, e.g., ultralytics reports)
3	Average IoU ≥ 0.7 on the validation set	0.5	Correct IoU implementation; Plot one image from the validation set and show predicted bounding boxes and ground truth
4	Clean code & inline documentation	0.5	Organized and readable

Part B (Take-Home — 3%)

#	**Requirement**	**Value (%)**	**Notes (Expectations for Excellent)**
5	Full YOLO training completed on full dataset	0.5	Model trained successfully; results reproducible, model under 200 MB
6	Quantitative evaluation (IoU, precision, recall, mAP)	0.5	Correctly calculated and interpreted
7	Held out test set performance	0.8	Your saved model will be evaluated on a hidden test set
8	Visualization of detections on validation images	0.5	Clear and accurate bounding box plots
9	Report documentation	0.7	Short discussion on YOLO performance vs two-stage detectors; Clear reasoning on speed vs accuracy; Motivate your choice for learning rate, confidence, and NMS thresholds